# CS 5350/6350: Machine Learining Spring 2018

Homework 0 Solutions

Yulong Liang (u1143816)

January 19, 2018

## Basic Knowledge

1. [5 points] **Answer:**

$$p(A) = 1 - (\frac{1}{2})^{10} = \frac{1023}{1024}$$

2. [5 points] **Answer:**
According to the Venn Diagram,

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) \leq p(A) + p(B)$$

Only when $p(A \cap B) = 0$, i.e., event $A$ and event $B$ are disjoint (never happen at the same time), the equality holds.

3. [10 points] **Answer:**
According to the proof above,

$$\begin{aligned} p(\cup_{i=1}^{n} A_i) &= p(\cup_{i=1}^{n-1} A_i \cup A_n) && (1) \\ &\leq p(\cup_{i=1}^{n-1} A_i) + p(A_n) && (2) \\ &\leq p(\cup_{i=1}^{n-2} A_i) + p(A_{n-1}) + p(A_n) && (3) \\ &\leq p(\cup_{i=1}^{n-k} A_i) + p(A_{n-k+1}) + \ldots + p(A_n) && (4) \\ &\leq p(A_1) + p(A_2) + \ldots + p(A_n) && (5) \\ &= \sum_{i=1}^{n} p(A_i) && (6) \end{aligned}$$

Only when all the events $A_1, A_n$ are disjoint (never happen at the same time), the equality holds.

4. [5 points] **Answer:**
According to the axiom of conditional probability,

$$\begin{aligned} p(A \cap B \cap C) &= p[A \cap (B \cap C)] && (1) \\ &= p(A|B \cap C)p(B \cap C) && (2) \\ &= p(A|B \cap C)p(B|C)p(C) && (3) \end{aligned}$$

5. [20 points]

(a) [10 points]

i. **Answer:**

| $X$ | 0 | 1 |
|---|---|---|
| $p(X)$ | 3/10 | 7/10 |

| $Y$ | 0 | 1 |
|---|---|---|
| $p(Y)$ | 4/10 | 6/10 |

ii. **Answer:**

| $X$ | 0 | 1 |
|---|---|---|
| $p(X\|Y=0)$ | 1/4 | 3/4 |
| $p(X\|Y=1)$ | 1/3 | 2/3 |

| $Y$ | 0 | 1 |
|---|---|---|
| $p(Y\|X=0)$ | 1/3 | 2/3 |
| $p(Y\|X=1)$ | 3/7 | 4/7 |

iii. **Answer:**

$$\mathbb{E}(X) = 1 \cdot p(X=1) + 0 \cdot p(X=0) = 0.7$$
$$\mathbb{E}(Y) = 1 \cdot p(Y=1) + 0 \cdot p(Y=0) = 0.6$$
$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 0.7 - 0.7^2 = 0.21$$
$$\mathbb{V}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = 0.6 - 0.6^2 = 0.24$$

iv. **Answer:**

$$\mathbb{E}(Y|X=0) = 1 \cdot p(Y=1|X=0) + 0 \cdot p(Y=0|X=0) = \frac{2}{3}$$
$$\mathbb{E}(Y|X=1) = 1 \cdot p(Y=1|X=1) + 0 \cdot p(Y=0|X=1) = \frac{4}{7}$$
$$\mathbb{V}(Y|X=0) = \frac{2}{3}(1-\frac{2}{3})^2 + \frac{1}{3}(0-\frac{2}{3})^2 = \frac{2}{9}$$
$$\mathbb{V}(Y|X=1) = \frac{4}{7}(1-\frac{4}{7})^2 + \frac{3}{7}(0-\frac{4}{7})^2 = \frac{12}{49}$$

v. **Answer:**

$$Cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0.4 - 0.7 \times 0.6 = -0.02$$

(b) [5 points] **Answer:**

$$p(X \cap Y) = 0.4 \tag{1}$$
$$p(X) \cdot p(Y) = 0.7 \times 0.6 = 0.42 \tag{2}$$
$$p(X \cap Y) \neq p(X) \cdot p(Y) \tag{3}$$

Thus, event X and event Y are not independent.

(c) [5 points] **Answer:**

$$\mathbb{E}(Y|X) = \begin{cases} \frac{2}{3}, & if\ X = 0 \\ \frac{4}{7}, & if\ X = 1 \end{cases}$$

$$\mathbb{E}(X|Y) = \begin{cases} 0.4 & , if\, Y = 0 \\ 0.75 & , if\, Y = 1 \end{cases}$$

Thus, they are not constant.

6. [10 points] **Answer:**
According to lognormal distribution,

(a) $\mathbb{E}(Y)$

$$\mathbb{E}(Y) = exp(\mu + \frac{\sigma^2}{2}) = \sqrt{e}$$

(b) $\mathbb{V}(Y)$

$$\mathbb{V}(Y) = exp(2\mu + \sigma^2) \cdot (exp(\sigma^2) - 1) = e \cdot (e - 1) = e^2 - e$$

7. [10 points] **Answer:**

(a) $\mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(Y)$

$$\begin{align} \mathbb{E}(\mathbb{E}(Y|X)) &= \int_{-\infty}^{\infty} \mathbb{E}(Y|X = x) f_X(x) dx \tag{1} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{Y|X}(y|x) f_X(x) dy dx \tag{2} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y \frac{f_{X,Y}(x,y)}{f_X(x)} f_X(x) dy dx \tag{3} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{X,Y}(x,y) dy dx \tag{4} \\ &= \int_{-\infty}^{\infty} y f_Y(y) dy \tag{5} \\ &= \mathbb{E}(Y) \tag{6} \end{align}$$

(b) $\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X))$

$$\begin{align} & \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)) \tag{1} \\ =& \mathbb{E}[\mathbb{E}(Y^2|X) - \mathbb{E}(Y|X)^2)] + \mathbb{E}[\mathbb{E}(Y|X)^2] - \mathbb{E}[\mathbb{E}(Y|X)]^2 \tag{2} \\ =& \mathbb{E}[\mathbb{E}(Y^2|X)] - \mathbb{E}[\mathbb{E}(Y|X)^2] + \mathbb{E}[\mathbb{E}(Y|X)^2] - \mathbb{E}[\mathbb{E}(Y|X)]^2 \tag{3} \\ =& \mathbb{E}(Y^2) - \mathbb{E}[\mathbb{E}(Y|X)^2] + \mathbb{E}[\mathbb{E}(Y|X)^2] - \mathbb{E}(Y)^2 \tag{4} \\ =& \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 \tag{5} \\ =& \mathbb{V}(Y) \tag{6} \end{align}$$

8. [20 points] **Answer:**
According to the context, the total number of heads c(n) is binomial distrubuted with number of trials equal to n and probability equal to 0.3, i.e., $c(n) \sim B(n, 0.3)$.

   (a) $\mathbb{E}(c(1))$, $\mathbb{V}(c(1))$
$$\mathbb{E}(c(1)) = n \cdot p = 1 \cdot 0.3 = 0.3$$
$$\mathbb{V}(c(1)) = n \cdot p \cdot (1 - p) = 1 \cdot 0.3 \cdot (1 - 0.3) = 0.21$$

   (b) $\mathbb{E}(c(10))$, $\mathbb{V}(c(10))$
$$\mathbb{E}(c(10)) = n \cdot p = 10 \cdot 0.3 = 3$$
$$\mathbb{V}(c(1)) = n \cdot p \cdot (1 - p) = 10 \cdot 0.3 \cdot (1 - 0.3) = 2.1$$

   (c) $\mathbb{E}(c(n))$, $\mathbb{V}(c(n))$
$$\mathbb{E}(c(n)) = n \cdot p = 0.3n$$
$$\mathbb{V}(c(n)) = n \cdot p \cdot (1 - p) = 0.21n$$

The expectation and variance of binomial distribution are propotional to the number of trials, namely $\mu = n \cdot p$, $\sigma^2 = n \cdot p \cdot (1 - p)$

9. [10 points] **Answer:**

   (a) $\nabla f(\mathbf{x})$
$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f(x)}{\partial x_1} \\ \dfrac{\partial f(x)}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(x)}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \dfrac{e^{-\boldsymbol{ax}} \cdot a_1}{(1 + e^{-\boldsymbol{ax}})^2} \\ \dfrac{e^{-\boldsymbol{ax}} \cdot a_2}{(1 + e^{-\boldsymbol{ax}})^2} \\ \vdots \\ \dfrac{e^{-\boldsymbol{ax}} \cdot a_n}{(1 + e^{-\boldsymbol{ax}})^2} \end{bmatrix}$$

(b) $\nabla^2 f(\mathbf{x})$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} \tag{1}$$

$$= \begin{bmatrix} \frac{a_1 a_1 e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} & \frac{a_1 a_2 e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} & \cdots & \frac{a_1 a_n e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} \\ \frac{a_2 a_1 e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} & \frac{a_2 a_2 e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} & \cdots & \frac{a_2 a_n e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{a_n a_1 e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} & \frac{a_n a_2 e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} & \cdots & \frac{a_n a_n e^{-\boldsymbol{ax}}(e^{-\boldsymbol{ax}} - 1)}{(1 + e^{-\boldsymbol{ax}})^3} \end{bmatrix} \tag{2}$$

(c) $\nabla f(\mathbf{x})$ when $\mathbf{a} = [1, 1, 1, 1, 1]^\top$ and $\mathbf{x} = [0, 0, 0, 0, 0]^\top$

$$\nabla f(x) = \begin{bmatrix} \frac{e^0 \cdot 1}{(1 + e^0)^2} \\ \frac{e^0 \cdot 1}{(1 + e^0)^2} \\ \frac{e^0 \cdot 1}{(1 + e^0)^2} \\ \frac{e^0 \cdot 1}{(1 + e^0)^2} \\ \frac{e^0 \cdot 1}{(1 + e^0)^2} \end{bmatrix} = \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

(d) $\nabla^2 f(\mathbf{x})$ when $\mathbf{a} = [1,1,1,1,1]^\top$ and $\mathbf{x} = [0,0,0,0,0]^\top$

$$\nabla^2 f(x) = \begin{bmatrix} \frac{a_1a_1e^0(e^0-1)}{(1+e^0)^3} & \frac{a_1a_2e^0(e^0-1)}{(1+e^0)^3} & \cdots & \frac{a_1a_5e^0(e^0-1)}{(1+e^0)^3} \\ \frac{a_2a_1e^0(e^0-1)}{(1+e^0)^3} & \frac{a_2a_2e^0(e^0-1)}{(1+e^0)^3} & \cdots & \frac{a_2a_5e^0(e^0-1)}{(1+e^0)^3} \\ \frac{a_3a_1e^0(e^0-1)}{(1+e^0)^3} & \frac{a_3a_2e^0(e^0-1)}{(1+e^0)^3} & \cdots & \frac{a_3a_5e^0(e^0-1)}{(1+e^0)^3} \\ \frac{a_4a_1e^0(e^0-1)}{(1+e^0)^3} & \frac{a_4a_2e^0(e^0-1)}{(1+e^0)^3} & \cdots & \frac{a_4a_5e^0(e^0-1)}{(1+e^0)^3} \\ \frac{a_5a_1e^0(e^0-1)}{(1+e^0)^3} & \frac{a_5a_2e^0(e^0-1)}{(1+e^0)^3} & \cdots & \frac{a_5a_5e^0(e^0-1)}{(1+e^0)^3} \end{bmatrix} \tag{1}$$

$$= \begin{bmatrix} 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \\ 0&0&0&0&0 \end{bmatrix} \tag{2}$$

10. [5 points] **Answer:**
Let $f(x) = e^x - (1+x)$, then $f'(x) = e^x - 1$, $f''(x) = e^x$. Since $f''(x) = e^x > 0$, $f'(x)$ increase for all $x \in \mathbb{R}$. When $x = 0$, $f'(x) = e^0 - 1 = 0$. Hence,

$$\begin{cases} f'(x) < 0 \ , x \in (-\infty, 0) \\ f'(x) = 0 \ , x = 0 \\ f'(x) > 0 \ , x \in (0, +\infty) \end{cases}$$

From the statement above, $f(x)$ is a convex function which reaches its local minimum at $x = 0$, $f_{min}(x) = f(0) = e^0 - (1+0) = 0$. Thus, $f(x) = e^x - (1+x) \geq 0$, i.e., $1 + x \leq e^x$.

Let $g(x) = e^{-x} - (1-x)$, then $g'(x) = -e^{-x} + 1$, $g''(x) = e^{-x}$. Since $g''(x) = e^{-x} > 0$, $g'(x)$ increase for all $x \in \mathbb{R}$. When $x = 0$, $g'(x) = -e^0 - 1 = 0$. Hence,

$$\begin{cases} g'(x) < 0 \ , x \in (-\infty, 0) \\ g'(x) = 0 \ , x = 0 \\ g'(x) > 0 \ , x \in (0, +\infty) \end{cases}$$

From the statement above, $g(x)$ is a convex function which reaches its local minimum at $x = 0$, $g_{min}(x) = g(0) = e^0 - (1-0) = 0$. Thus, $g(x) = e^{-x} - (1-x) \geq 0$, i.e., $1 - x \leq e^{-x}$.